



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Book Review of: Self-Knowledge, Edited by Hatzimoysis, A. Oxford University Press

Citation for published version:

McGlynn, A 2011, 'Book Review of: *Self-Knowledge*, Edited by Hatzimoysis, A. Oxford University Press', *Notre Dame Philosophical Reviews*. <<http://ndpr.nd.edu/news/25450-self-knowledge/>>

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Notre Dame Philosophical Reviews

Publisher Rights Statement:

© McGlynn, A. (2011). Book Review of: *Self-Knowledge*, Edited by Hatzimoysis, A. Oxford University Press. Notre Dame Philosophical Reviews.

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Anthony Hatzimoysis (ed.), *Self-Knowledge*, Oxford University Press, 2011, 285pp., \$85.00 (hbk), ISBN 9780199590728

Reviewed by Aidan McGlynn, the Northern Institute of Philosophy at the University of Aberdeen

What are philosophers of self-knowledge aiming to do? Characterize the nature and scope of so-called privileged access (henceforth PA), and account for its distinctive character? Explain (or explain away) the peculiar security that we enjoy when avowing our own mental states, in contrast with when we issue reports about how things stand with other contingent matters? Show once and for all that (whatever it amounts to) PA is in fact compatible/incompatible with content externalism, of the sort familiar from the writings of Putnam and Burge? Account for our ability to rationally police our own attitudes? Prove that virtually everything Descartes wrote about the mind is wrong? Throw light on the puzzling features of first person thought and talk, such as Moore's paradox, apparent guaranteed referential success, and immunity to error through misidentification?

The only answer that would prompt anything like widespread assent is probably 'some of the above', the usual rules of the game being that, whatever it is that one does, one needs to do it while avoiding the aspects of the Cartesian picture that prompt worries about our knowledge of other minds, private languages, and the like. The thirteen papers in this volume (written by thirteen distinguished philosophers, whose primary interests are spread across epistemology and the philosophies of mind and language) attempt to do, well, some of the above. Here I'll pick my way through some of the contributions, focusing on those about which I've found myself with the most to say, and summarizing the rest in my concluding remarks. I should emphasize, though, that some of the papers skipped over here are amongst the most interesting in the volume.

The Scope of PA

In the opening paper, Ram Neta poses and attempts to answer an important and rather neglected question. It's widely assumed that we enjoy PA to our beliefs, hopes, sensations, and so on, but not to our 'knowings', 'seeings' (in the object-involving sense) or 'rememberings', even though a case can be made that the latter are just as much aspects of our mental lives as the former. But what (if any) reason do we have for thinking that the reach of PA is limited in this fashion?

Neta's answer comes in two parts. First, he argues for a *truth-sufficiency* account of PA, according to which what is distinctive of a fact to which one enjoys PA is that it justifies one in believing that very fact (19).¹ Second, he argues that it's a necessary condition on S having a justification q to believe p that S be able to believe p when and because of q (28). Putting these together, we have it that one has PA to the fact

¹ Page numbers unaccompanied by further references are to the volume under review.

that q just when one has the ability to believe that q when and because of q (29). While we have this ability with respect to the relevant facts about our beliefs, hopes, and sensations, we lack it with respect to facts about our knowings, seeings, and rememberings. That's why PA gives out where it does.

I have several misgivings with this proposal. First, the truth-sufficiency account seems under-motivated and problematic. By way of motivation, Neta suggests (21) that it explains the phenomenon of immunity to error through misidentification (henceforth IEM) and related phenomena. But this is implausible, even before we see the details. As Evans and subsequent commentators have stressed, the class of judgments that display IEM outstrips the class of judgments that exploit PA (an example is offered below), and so the proposed explanation is at best incomplete. As for the truth-sufficiency account itself, it only offers an account of PA to propositional attitudes if such attitudes can be presented to consciousness in such a way as to provide a reason or justification for the belief that one has those attitudes. We have, at present at least, no promising model of how this might work (Coliva 2008). Finally, those (myself included) who take the apparent *groundlessness* or *baselessness* of basic self-knowledge more seriously than Neta does (22) will regard the truth-sufficiency account's obvious compromise of this as further cause for concern.

In any case, Neta doesn't really offer any reason for thinking that the distribution of relevant abilities is as the standard picture of the reach of PA needs it to be, given his proposal. He's well aware of this, and seems inclined not to worry, suggesting that we should be open to the possibility that further investigation may reveal the reach of PA to be rather more limited or more extensive than we typically assume (30-1). But if we already have reservations about Neta's proposal, this is liable to seem unsatisfying.

Still, while I'm sceptical of Neta's own response, he raises a significant challenge. Many of us will want to hold that one can have PA to what one believes even if content externalism is true, but Neta worries that reconciling these (in the manner developed further by Crispin Wright in his contribution to this volume, for example) may undermine the obvious motivation for holding that one cannot have PA to one's knowings, rememberings, and seeings; that if we could, we could put that together with our knowledge—presumably also available from the armchair—of the factivity of knowing and remembering, and the object-involving nature of the relevant notion of seeing, in order to yield armchair knowledge of contingent facts about the external world. The matter deserves more careful consideration than I can give it here.

Sven Bernecker's contribution argues that the scope of PA is narrower than philosophers are prone to supposing. He makes a case that the representationalist theory of mind, familiar from the work of Fred Dretske and others, entails that while one can have armchair knowledge that one knows p (rather than q), one cannot have armchair knowledge that one knows p (rather than having no knowledge at all).

In order to give this conclusion a bit more bite, Bernecker starts out by trying to motivate the claim that we *are* able to have the kind of second-order armchair knowledge denied by his conclusion.² However, this motivation consists of a very unpersuasive defence of the KK principle (or rather, a defence of the claim that KK is ‘not *obviously* false’ (37), since there’s something to be said in its favour, and familiar objections to it aren’t as decisive as sometimes thought). The only consideration Bernecker offers on behalf of KK is that, taken together with the knowledge norm of assertion (henceforth the K-norm), it offers an explanation of the absurdity of assertions of the form ‘p but I don’t know p’—what I’ll call *epistemic Moorean assertions*. By the K-norm, my asserting some proposition q implies that I know q. By KK, my knowing q implies that I’m in a position to know that I know q. So it follows that my asserting q implies that I’m in a position to know that I know q.³ Returning now to our Moorean conjunction, my asserting p implies that I’m in a position to know that I know p, while the second conjunct denies that I know this—small wonder my assertion of the conjunction sounds absurd.

But all this is liable to impress as relying on a pun on ‘implies’. Given the K-norm, asserting q implies one knows q in the sense—the one operative in Moore’s original discussion—that one *represents oneself* as knowing q. In contrast, my knowing q *entails* that I’m in a position to know that I know q, according to KK. It doesn’t follow that in asserting P I represent myself as being in a position to know that I know P, since neither my audience nor I may have the slightest inkling that KK holds. In any case, Bernecker doesn’t consider any rival explanations of the absurdity of epistemic Moorean assertions; in particular, he doesn’t consider the better known, more economical, and less problematic explanation that dispenses with KK and just appeals to the K-norm (see Williamson 2000).

So I’m inclined to think that if the KK principle proves to be in tension with a well-motivation conclusion, we can surrender KK without much fallout. That said, I have reservations about whether Bernecker’s conclusion *is* well-motivated. His principal argument starts by assuming without argument⁴ that knowledge (or ‘knowledge’) is somehow sensitive to context (where this is to be interpreted so as to be neutral between contextualists proper and their shifty-invariantist rivals), and by ‘spell[ing] out’ (41) this supposed insight with a contrastivist version of the familiar relevant-alternatives picture. According to contrastivism, ‘S knows that p’ expresses a three-place relation between a subject, a proposition p, and a contrast proposition q (which can shift with shifts in context). This allows Bernecker to distinguish the question of whether one can have armchair knowledge that one knows p rather

² Bernecker also tries to make the conclusion of his argument more puzzling by suggesting that if it’s true, there’s no warrant to be had for one of the premises of that argument (45-6). I can’t help but suspect that this instability makes the argument less interesting, not more, though I readily concede that the issue is not straightforward.

³ The text concludes at this point that when one asserts q, one is in a position to know that one knows q (36). I assume that this is a typo.

⁴ Bernecker just points out that some sort of sensitivity to context is ‘widely assumed’ (40).

than that one knows q from the question of whether one can have armchair knowledge that one knows p rather than that one lacks any knowledge whatsoever. Bernecker's central claim, then, is that representationalism entails that one can have the former but not the latter. But the significance of this conclusion was rather lost on me, given my uncertainties and concerns about the path by which Bernecker reaches it.

Neo-Expressivism

Anthony Brueckner contributes a critical notice of Dorit Bar-On's excellent 2004 book, which offered the most sophisticated and detailed version of an expressivist account of the distinctive security of avowals yet produced, and Bar-On is given a chance to clarify her position and respond to Brueckner's worries. Traditional expressivism goes wrong, according to Bar-On, in failing to distinguish various notions of expression. The sentence 'I have a headache' *semantically expresses* (s-expresses) the boring, truth-conditional content you learned it does at your mother's knee, even when avowed—this largely accounts for the 'neo' in the account's name. But Bar-On holds that in issuing that avowal, one also intentionally does something to directly express *one's headache*: one *expresses in the action sense* (a-expresses), as Bar-on, following Sellars, puts it (see 177 and 194 for the distinction). Moreover, Bar-on also holds that one also a-expresses the belief that one is in a headache. She calls this the *dual-expression thesis*, though a perhaps more helpful name would be the *dual-a-expression thesis*. The 'neo' bit of the view allows Bar-On to escape the usual Frege-Geach-style worries with expressivism (allows her to preserve *semantic innocence*, as she prefers to say), while the 'expressivism' bit enables her to explain the distinctive features of avowals—that they're so strongly presumed to be true, aren't subjected to the usual epistemic assessment, and so on.

In this exchange, Brueckner lobbs about eleven objections at neo-expressivism, and Bar-On plays defence. Each ends up with a rather mixed scorecard. In my opinion, Bar-On successfully sees off some of Brueckner's worries. For example, I think Bar-On is entirely correct to insist, contrary to Bruckner's suggestion (178), that a-expression is not at bottom a causal notion. And I think she makes a good case that some of Brueckner's objections turn on misunderstandings of her view.

Others, however, remain. Like Brueckner (175-6), I'm inclined to think that there's no insight that can be gleaned from Burge's discussion of 'self-verifying' judgments (such as that expressed by 'I am thinking that water is wet') and generalized to avowals as a class. I didn't understand Bar-On's reply to this worry (191-3). Her suggestion seems to be that while Burge recognizes the need to get away from a model on which knowing that one is thinking that water is wet requires bringing recognitional capacities to bear to determine the content of one's thought, he has nothing to put in place of that model. Expressivism is needed to fill the gap. But, like Brueckner, I can't see any such gap in Burge's treatment of self-verifying judgments. And despite Bar-On's responses, I continue to share Brueckner concerns that her account of privileged self-knowledge rests on a very radical and implausible version

of disjunctivism (184-5), and that her account of the security of avowals seems to push the question back in an unsatisfactory way (180).

Moreover, Bar-On doesn't so much as mention the biggest problem raised by Brueckner. In Bar-On's book she discussed the challenge posed to neo-expressivism by 'negative avowals' (such as 'I don't want coffee', 'I'm not in intense pain', or 'I don't believe that the number of stars is odd'). We might say that one is expressing the desire *not* to have coffee in the first case, but this doesn't readily generalize to the others. Given central features of her account, Bar-On cannot say that one expresses the absence of the relevant desire, pain, or belief, or that in each case one expresses some more complex mental state (as Mark Schroeder proposes on behalf of ethical expressivism as the best response to a structurally similar problem concerning what ethical judgments s-express).⁵

In Bar-On's discussion of this problem in her book she claims that so-called negative avowals aren't really avowals; rather, one introspects, and then reports not finding the relevant mental state, such reports being 'open to doubts, questioning, requests for reasons/explanation' (2004: 335). But this won't do. As Brueckner points out (184-5), some negative 'avowals' (for instance, 'I am not in intense pain') seem just as secure, in the relevant sense, as most positive avowals. We may add that Bar-On simply isn't in a position to appeal to introspection here. She argues against introspectionist accounts of the security of (positive) avowals on the grounds that propositional attitudes don't have any distinctive phenomenology for one to introspect (2004: 103), and she also argues that, contrary to the traditional picture, there's no interesting sense in which one's mental states are transparent or salient to one (2004: 406-8). Given this, one shouldn't have *any* confidence that failing to introspectively locate the belief that the number of stars is odd signifies that one does not have that belief. If negative avowals are reports of failures to introspect some mental state, Bar-On should maintain that we have no reason to trust them at all, since we cannot rule out false-negatives being the rule rather than the exception. Negative avowals continue to pose a threat to Bar-On's view, and it's a shame that she didn't take the opportunity to revisit the topic here.

The First Person and IEM

José Luis Bermúdez's admirably well-written and enjoyable paper draws from the deep well provided by *The Varieties of Reference* in order to try to explain the connections between self-knowledge and the first person. Bermúdez first argues for a plausible constraint he calls the *symmetry constraint*, according to which any account of the sense of "I" must 'preserve the possible token-equivalence of "I" and other personal pronouns with respect to same-saying and relative to a particular context' (229), and draws from this the conclusion that we need to adopt a distinction between what he calls the token-sense of "I" and its type-sense. The former is 'what a speaker or hearer understands when they utter or hear an

⁵ Thanks to Aaron Cotnoir for pointing out the similarity here.

utterance of “I”, while the latter is ‘that in virtue of which a speaker or hearer can properly be said to understand the expression “I” (231). The distinction is needed since “you” and “I” clearly do not have the same type-sense even though (by the symmetry constraint) the token-sense of “I” can be the same as the token-sense of “you” in a given context.

Bermúdez convincingly argues that, contrary to an initially tempting thought, none of the distinctions made in either Kaplan’s or Perry’s treatments of indexicals captures the distinction he wants to draw between the token-sense of “I” and its type-sense, and that Evan’s account of the sense of “I”, which appeals crucially to our sensitivity to sources of judgments that display IEM, is misplaced. Instead, Bermúdez appeals to a version of another of Evans’s ideas, namely that to understand an utterance of “I” in a particular context requires being able to locate the speaker in the objective spatiotemporal order. However, Bermúdez makes two important adjustments. First, while Evans saw this ability as a further component of the sense of “I”, supplementing the other components involving IEM, Bermúdez thinks it stands by itself as ‘all that we need in order to understand the token-sense of “I” (241). And second, Bermúdez thins-down what is required to have this ability in order to make it more plausible that the resulting account meets the symmetry constraint. As Mark Sainsbury nicely puts the proposal in his reply, ‘the ability to know the speaker’s location can be the common element in an account both of her capacity for self-reference and of another’s capacity for understanding’ (248). That this ability can be a ‘common element’ here is what makes it suitable to feature in an account of the token-sense of “I” that respects the symmetry constraint.

This proposal retains something of the spirit of Evans’s account while being decidedly more plausible. But I found myself sympathetic to the criticisms developed in Sainsbury’s reply. Sainsbury argues that even Bermúdez’s thinned-down constraint on understanding an utterance of “I” is implausibly demanding, at least if it’s to be understood in such a way that it has any teeth at all (249-52), and that knowing who made an utterance involving “I” is the wrong kind of condition to impose on understanding such an utterance, since what it takes to satisfy it is much too sensitive to context (253-4).

Sainsbury’s own positive proposal is that ‘there’s no more to understanding a token of “I”, whether as speaker or hearer, than being able to apply to the token the rule: English speakers should use “I” to refer to themselves as themselves’ (254). I won’t dispute this claim as a piece of semantic epistemology (to borrow a useful phrase from Bar-On 2004). I am, however, sceptical of Sainsbury’s claim that the account explains why some items of self-knowledge display IEM (258-9). First, I’m worried the explanation has the same defect as Neta’s, discussed above; given the range of judgments that can be IEM, it’s *prima facie* implausible that the IEM of certain “I”-judgments is to be explained by a rule that’s distinctive of that pronoun.

More importantly, Sainsbury sees IEM as a form of guaranteed referential success, while I want to insist that these are two distinct phenomena that may need to be

handled rather differently (see the early chapters of Bar-On 2004 for useful discussion). To take a familiar example, my judgment ‘That is moving really fast!’, made on the basis of perception, can be IEM; we can readily make sense of the idea that I’m wrong about whether the object I’m demonstrating is moving fast, but there’s no sense to be made of the suggestion that I’m right that *something* is moving fast, but I’ve misidentified it as *that*. The latter sort of error hasn’t been provided for, to borrow Wittgenstein’s well-worn phrase. But my uses of *that* as a visual demonstrative aren’t guaranteed to refer to anything. It’s plausible enough that Sainsbury’s account of the rule governing “I” explains why my uses of “I” to refer to myself seem more or less guaranteed to pick out the right object; I cannot follow the rule without thinking of myself as myself (259). But I think it’s a mistake to think that this throws any light on the phenomenon of IEM. I suspect that the correct account of IEM will, as Sainsbury writes of other phenomena, ‘need to be found elsewhere than in the nature of self-reference’ (260).⁶

Concluding Remarks

Elsewhere in the volume: Gary Ebbs offers a remarkably complex new argument against what he calls ‘the Standard Analysis of Epistemic Possibility’; Crispin Wright refines the notion of transmission failure, and modifies and qualifies his earlier diagnosis of the McKinsey paradox; Alex Byrne argues that the transparency method—suggested by Gareth Evans, and developed as an epistemology of belief, intention, desire and so on, in Byrne’s other writings on self-knowledge—can also yield knowledge of what one is thinking; Brie Gertler does a good job of articulating some of the reasons to regard Byrne’s account, and transparency accounts more generally, as implausible; André Gallois offers a difficult but interesting discussion of the interrelations between Moore’s paradox, self-knowledge, and *de se* knowledge, one that appeals to a slightly different interpretation of the transparency method to that found in Byrne’s work; Charles Travis turns in one of the most opaque pieces of philosophical writing that I have encountered recently⁷; and finally, David Owens offers an excellent discussion of the relationship between first person thought and ‘mental self-control’, as it bears on Descartes’s project in the *Meditations*.

It must be said that this volume wouldn’t make a particularly good introduction to the topic of self-knowledge. Earlier proposals are revisited, theories are applied to

⁶ Sainsbury and Tye 2011 and forthcoming supersede various aspects of Sainsbury’s position in the present paper, though they are continuous in maintaining that there’s no distinctive kind of *content* grasped when one thinks of oneself as oneself. The latter work revisits IEM and guaranteed referential success, and distinguishes them in much the way I have in the text above.

⁷ A quotation from the opening page of the paper (202), lest this come across as unduly dismissive: ‘There is a way for someone’s mental life to be—for *anyone* to think it—only where *one* may think that life to be that way; so only where *one* can think of the one whose life it is that his being as he is is his being that way; so only where *one* can get in mind what *is* that person being that way. What, then, might mental life be? Specifically, what might visual experience be? That is our topic.’ Reading twenty-four pages written like this is not an enjoyable experience, though it appears to be the result of a stylistic choice on Travis’s part. I cannot see any justification for making life quite so difficult for one’s audience.

new cases, principles are Chisholmed, objections to long-standing views are refined or rebutted, and so on. In this respect, many (though by no means all) of the papers have a starting point that is already quite embedded in recent debates about these issues.⁸ For those already up to their pineal glands in these debates, the volume has considerably more to offer. There is much to be learned from the attempts of these philosophers to wrestle with the issues and problems associated with the seemingly distinctive character of self-knowledge and first person thought and talk: it goes almost without saying that there is much to disagree with and puzzle over, but equally there is much to admire and agree with. One thing is for sure; one will take away a renewed sense of just how difficult these issues really are.⁹

References

Bar-On, D. 2004. *Speaking My Mind: Expression and Self-Knowledge*. Oxford University Press.

Coliva, A. 2008. Peacocke's Self-Knowledge. *Ratio* 21: 13-27.

Gertler, B. 2010. *Self-Knowledge*. Routledge.

Sainsbury, R. M., and Tye, M. 2011. An Originalist Theory of Concepts. *Proceedings of the Aristotelian Society Supplementary Volume* 85: 101-24.

Sainsbury, R. M., and Tye, M. Forthcoming. *Seven Puzzles of Thought and How to Solve Them: An Originalist Theory of Concepts*. Oxford University Press.

Williamson, T. 2000. *Knowledge and its Limits*. Oxford University Press.

⁸ Gertler 2010 offers an up-to-date and clear introduction to many facets of these debates.

⁹ My thanks to the members of the Self-Knowledge pilot project at the Northern Institute of Philosophy, and to Douglas Edwards and Mark Sainsbury.